

Durham Research Online

Deposited in DRO:

15 May 2018

Version of attached file:

Published Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Bentley, R. Alexander and Acerbi, Alberto and Ormerod, Paul and Lamos, Vasileios (2014) 'Books average previous decade of economic misery.', PLoS ONE., 9 (1). e83147.

Further information on publisher's website:

<https://doi.org/10.1371/journal.pone.0083147>

Publisher's copyright statement:

© 2014 Bentley et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

Books Average Previous Decade of Economic Misery

R. Alexander Bentley^{1*}, Alberto Acerbi¹, Paul Ormerod^{2,3}, Vasileios Lamos⁴

1 Department of Archaeology & Anthropology, University of Bristol, Bristol, United Kingdom, **2** Volterra Partners LLP, London, United Kingdom, **3** Anthropology Dept, Durham University, Durham, United Kingdom, **4** Department of Computer Science, University of Sheffield, Sheffield, United Kingdom

Abstract

For the 20th century since the Depression, we find a strong correlation between a ‘literary misery index’ derived from English language books and a moving average of the previous decade of the annual U.S. economic misery index, which is the sum of inflation and unemployment rates. We find a peak in the goodness of fit at 11 years for the moving average. The fit between the two misery indices holds when using different techniques to measure the literary misery index, and this fit is significantly better than other possible correlations with different emotion indices. To check the robustness of the results, we also analysed books written in German language and obtained very similar correlations with the German economic misery index. The results suggest that millions of books published every year average the authors’ shared economic experiences over the past decade.

Citation: Bentley RA, Acerbi A, Ormerod P, Lamos V (2014) Books Average Previous Decade of Economic Misery. PLoS ONE 9(1): e83147. doi:10.1371/journal.pone.0083147

Editor: Matjaž Perc, University of Maribor, Slovenia

Received: August 7, 2013; **Accepted:** October 30, 2013; **Published:** January 8, 2014

Copyright: © 2014 Bentley et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: A. Acerbi is supported by a Newton International Fellowship. V. Lamos acknowledges the support from the TrendMiner project (EU-FP7-ICT n.287863). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: R.A. Bentley is a PLOS ONE Editorial Board member and P. Ormerod is employed by a commercial company (Volterra Partners LLP). This does not alter the authors’ adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: r.a.bentley@bristol.ac.uk

Introduction

Whereas it is self-evident that individuals must rely on past observations to anticipate the consequences of future decisions [1–3], the time scale over which observations are aggregated, particularly at the population level, is a pertinent evolutionary question. A characteristic time-scale of about 30 years in the usage of new words, for example, may be a generational effect within recent centuries of language evolution [4]. This may be a relatively modern effect, however, and now the instantaneous nature of the Internet contrasts even more with millennia of cultural evolution that allowed humans to accumulate information and learn skills over many generations [5–9].

Over a much longer time scale, written language not only accumulates technological knowledge but it continually regenerates the cultural basis by which people make sense of their own experience and maintain their social relationships [10,11]. The weighting of this experiential knowledge can be affected through the expression of associated emotions [10,12–16].

We hypothesize that book authors, as both producers and consumers of shared knowledge, are informed by past economic conditions at some definable time scale – they convey common knowledge not just factually but also indirectly, through emotional content. Accordingly, we would expect to find a positive correlation between mood, as expressed in books, and economic conditions in the recent past. As books take time to write and publish, we expect that each year of publication may average past economic influences over a definable number of years.

The relationship between emotions and current events can be explored through the increasing availability of recorded language use at the population level [17–24]. Language expressing emotion has recently been correlated with socio-political and economic trends [25–28], amid a wealth of studies exploring how large

samples of online language use can be statistically predictive future consumer activity, unemployment rates, concerns about personal health and other collective behavior [29–35]. A “forward looking” index using Google trend data that shows strong correlation with GDP per capita further evidences the connection between crowd-sourced textual data and broad economic indicators [36].

To complement the vast study of social media on the scale of days or hours, we explore whether human sentiments, aggregated by language use at the population scale, reflect not just with specific historical events but with more general economic conditions of the past. Focussing on the 20th century, through the Google Books Ngram corpus [4,37–39], we examine how well a new index of emotions in books [40], or literary ‘misery index’, LM_t , correlates with the widely-used economic ‘misery index’, EM_t , which is the unemployment rate plus the inflation rate [41,42].

Our simple model is that the literary misery index will be proportional to the moving average of economic misery, $LM_t \propto EM(\tau)_t$, where τ denotes the period over which we take the moving average of the annual U.S. misery index scores (years leading up to and including the current year). The moving average is based solely on the past, in that it is a non-weighted arithmetic mean of the past (τ) years of economic misery.

As described in the Methods section below, we used the WordNet Affect (WNA) text analysis tool [43–45] to extract literary mood from the Ngram corpus [40]. We also compared our results with two independent emotion extraction tools, Linguistic Inquiry and Word Count [46] and a recently proposed “hedonometer” method [47].

Results

Figure 1 compares the time series of the literary misery index, LM_t , derived from the WNA sampling of all English books (see Methods), versus the contemporaneous U.S. economic misery index $EM(1)_t$. There is some correlation (Pearson's $r=0.25$) even when we do not take any moving average of the economic misery index, $EM(1)_t$. Visually, the literary misery index seems to respond to major phases of the 20th century: literary misery increased after the economic Depression, then declined after the post-War years, then rose again after the recession of the 1970s, and declined again following on from the economic recovery of the late 1980s. The literary misery time series is well-characterised ($r=0.711$) by a sine wave with a period of 41.6 years, phased such that a t_0 year would be 1890. Spectral analysis also indicates a similar periodicity in the economic misery index, though this is not as clearly defined, being of the order of 25–50 years. Importantly, this frequency range is far more important than the period of 5–12 years which is widely believed to characterise the economic business cycle of the fluctuation in real GDP growth.

When we compare the correlations between LM_t and all the values of moving average (τ up to 25 years) for the economic misery index (see Figure 2a), we find a peak in the overall goodness of fit at $\tau=11$ (Pearson's $r=0.54$). The moving average correlates significantly better than the best fit using a simple lag, which is $r=0.42$ at a lag of 7 years, i.e. between LM_t and economic misery 7 years previous. Figure 2b shows the time series for $EM(11)_t$, the 11-year moving average of the U.S. misery index, versus literary misery index, LM_t , derived from all books calculated through WNA. Notice that, as shown in Figure 1, the volatility of the economic conditions has been larger than have been the changes in literary misery, so LM_t has been scaled by a factor of 10 to allow a better comparison.

A closer relationship between LM_t and $EM(11)_t$ distinctly begins with the economic Depression following the 1929 crash (Figure 2b). When we compare the two time series considering only years from 1929 to 2000, we find again a peak in the overall goodness of fit at $\tau=11$ years (Pearson's $r=0.79$), and correlations are overall considerably better than for the entire 20th century, for all values of moving average τ up to 22 years (Figure 3a), and still significantly better than a lag (the lag peaks at $r=0.62$ for a lag of

8 years). The scatterplot in Figure 3b shows how the literary misery score, LM_t , correlates with the 11-year moving average of yearly U.S. Misery index, $EM(11)_t$, from 1929 to 2000.

It could be argued that the results are primarily due to three data points. Economic misery coincides with WW1 (1918), the aftermath of the Great Depression (1935) and the energy crisis (1975). We addressed this point directly by calculating the correlations on a 'leave-one-out' basis. In other words, when a correlation is calculated using data from $1, \dots, n$, leave-one-out calculate the correlation over the periods $(2, \dots, n; 1, 3, \dots, n; 1, 2, 4, \dots, n, \text{ etc.})$. The results are therefore robust, and do not depend upon the potential 'extreme' data points.

The economic misery index, $EM(\tau)_t$, correlates better with the literary misery index than do the individual components of the economic misery index – unemployment or inflation – on their own. As shown in Figure 3c, these correlations with U.S. misery index, $EM(\tau)_t$, are better than the correlations versus U.S. inflation (best $r=0.34$) or U.S. unemployment (best $r=0.36$). Note that while both correlate positively with literary misery, U.S. unemployment and U.S. inflation themselves were moderately anti-correlated during the 20th century ($r=-0.33$ for whole century and $r=-0.41$ from 1929 to 2000), such that their sum, economic misery, best correlates with literary misery via the moving average. Note that the U.S. and U.K. economy misery indices were themselves correlated from 1929 to 2000 ($r=0.65$).

Table 1 shows the statistically significant correlations between $EM(11)_t$ and literary misery indices calculated in other subsets of the English corpora, or in the same main corpus (all Books in English) but with other techniques (see Methods). We also calculated correlations between $EM(11)_t$ and WNA and LIWC single moods. Interestingly, all the best correlations involve our 'composite' literary misery index (as opposed to single moods). Two independent ways to calculate the literary misery index (LIWC and Hedonometer) give positive and significant correlations with $EM(11)_t$. Comparing all values of moving average for all English books up to $\tau=25$ years, the same value of $\tau=11$, which yielded the best correlation for WNA, also gives the best correlation both for LIWC and Hedonometer Table 1), while the very best correlations for British books ($r=0.47$) or American books ($r=0.66$) alone are at $\tau=10$.

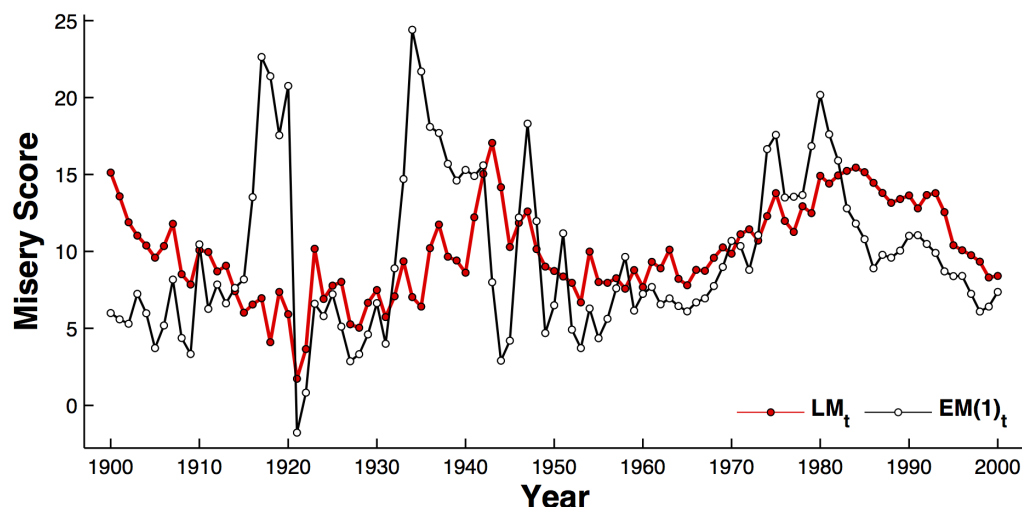


Figure 1. Time series of the literary misery index LM_t for all books calculated through WNA (white circles), versus the U.S. economic misery index $EM(1)_t$ (red circles). LM_t has been scaled by a factor of 10 to allow a better comparison. doi:10.1371/journal.pone.0083147.g001

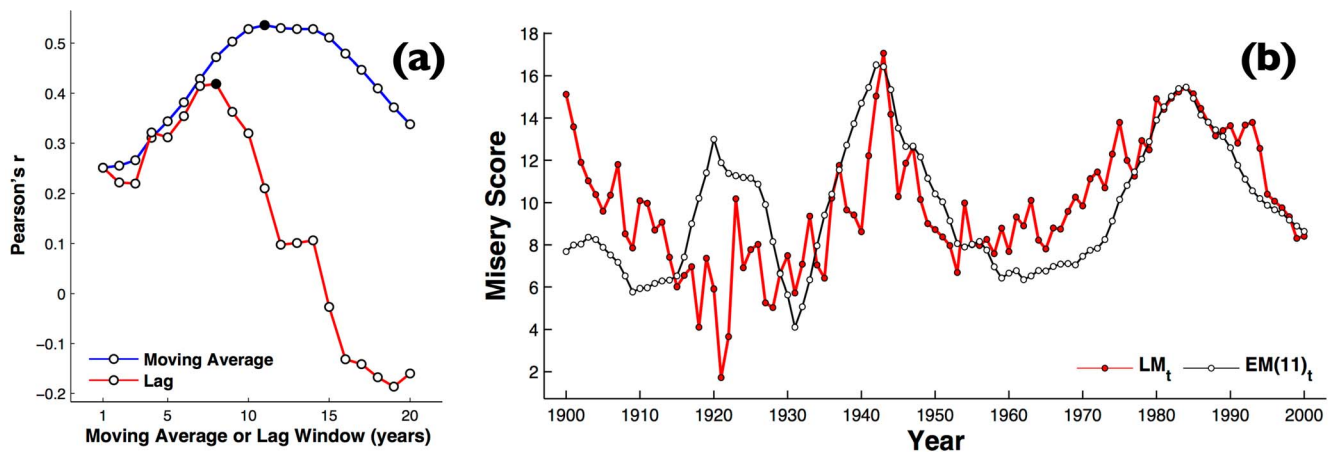


Figure 2. Moving average of economic misery. (a) The effect of varying the moving average period, τ , versus a simple lag, on correlation expressed as Pearson's r between the time series of LM_t for all books of 1900–2000 and U.S. misery index (b) Time series for $EM(11)_t$, the 11-year moving average of the U.S. misery index (red circles), versus literary misery index, LM_t , derived from all books calculated through WNA (white circles). Similarly to Figure 1, LM_t has been scaled by a factor of 10 to allow a better comparison. doi:10.1371/journal.pone.0083147.g002

These results are robust, but to explore other possibilities we repeated the analyses on a composite index in which we simply averaged the U.S. and British misery indices for each year, and again found that the best correlation ($r=0.82$) with WNA 'misery' extracted from All English Books occurred at a moving average of 11 years, and that the best correlation of the averaged UK/US economic misery with LIWC misery ($r=0.64$) occurred at 10 years moving average.

Finally, to check these patterns against a non-English language, we analysed the correlations of a literary misery index derived from LIWC sampling of German books, with the German EM index from 1929–2000. We get an optimal Pearson correlation of $r=0.60$ for a window of 10 years (Figure 4).

Discussion

We have found a distinct positive correlation between a literary misery index in English (and German) language books and a moving average of the economical misery index. It is a common observation that culture cycles [20,48–50], and given that both literary misery and economic misery follow cycles, it could be counter-argued that the correlation here has no causal link.

We think it is highly unlikely that there would be no causal relationship, for a few reasons. Firstly, the best correlation is achieved when we match literary misery with economic misery specifically, rather than different emotions compared to economic misery, or literary misery to unemployment or inflation individually. This implies a causal connection; it actually matters that we match the emotional index to the economic index.

Secondly, the moving average of economic misery correlates better with literary misery than economic misery after a specific number of years time lag. The results support the hypothesis that emotion words in books reflect general past economic conditions averaged over a time scale of approximately the previous decade. This underlies the logic of causal connection we are proposing, which is that populations accumulate experience over a period of time, rather than referring back to a period of some specific number of years ago, via a lag.

Finally, there is sociological reality that unemployment, one of the two components of economic misery, generally has a negative effect on emotional state [51–53]. Unemployed workers are more

likely to experience depression, anxiety, and poor self-esteem [54] and even increased mortality rates related to alcohol and smoking [55,56]. This affects not just individuals, but also families and communities [57,58].

At the population scale, it seems to take about a decade for these effects to become registered in books. Using different techniques for calculating the literary misery index on multiple corpuses, we find agreement for the best period of 10–11 years for the moving average. This is somewhat shorter than the generational time scale observed in the usage of new words [4] or in the collective drift in 'content-free' words tied to group style and identity [59]. Underlying this difference may be the event-driven nature of economics, injecting a continual stream of novelty into cultural evolution that is perhaps more subject to the limits of individual memory retrieval [1,60].

Perhaps this 'decade effect' reflects the gap between childhood, when strong memories are formed, and early adulthood, when authors may begin writing books. Consider, for example, the dramatic increase of literary misery in the 1940s, which follows the Great Depression of the 1930s. As documented by a long term study of 167 Californians born in 1920–1921, the Depression Era left long-lasting impressions on their memories [61]. Authors from this generation would have begun writing in the 1940s.

In fact, the most significant correlation between emotions in books and past economics begins with the year 1929, persisting for the remainder of the 20th century. This suggests that the global Depression instigated a more direct relationship between books and past economics. New urbanisation, mass production, and advertising in the 1920s appear to have effected a lasting change in emotional expression in literature [62]. The subsequent increase in 'narcissistic' word usage [26,28] may reflect a growing 20th century trend to interpret global economic events on a personal emotional level.

Written language enables the accumulation of factual knowledge, an evolutionary process as old as Palaeolithic cave art [8,63]. In smaller societies, shared emotional knowledge provides an incentive and means for humans to cooperate [64]. One of these means is indirect speech, which incorporates shared knowledge of other people's feelings, and helps preserve social relationships by providing ways of 'saving face' amid continual complex negotiations of cooperation [11]. Joking relationships, for example, which

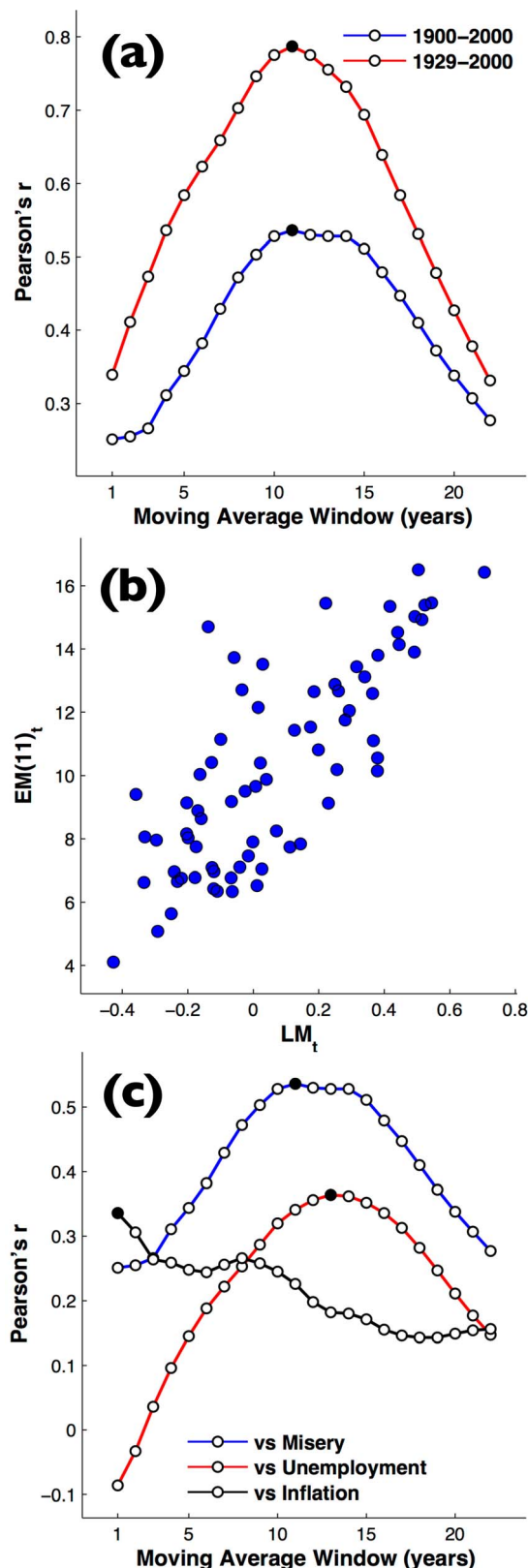


Figure 3. Time period and economic index. (a) The effect of varying the moving average period, τ , on correlation expressed as Pearson's r between the time series of LM_t for all books in the Google data and U.S. economy misery index, $EM_t(\tau)$ from 1900 to 2000 (blue line) and between the the two same time series from 1929 to 2000 (red

line) (b) Scatterplot of literary misery index, LM_t , versus the 11-year moving average of U.S. economy misery index, $EM(11)_t$, from 1929 to 2000. (c) The effect of varying the moving average period, τ , on correlation expressed as Pearson's r between the time series of LM_t for all books in the Google data of 1900–2000 versus U.S. misery index (blue line), U.S. unemployment rate (red line), and U.S. inflation rate (black line).

doi:10.1371/journal.pone.0083147.g003

require shared emotional and kinship knowledge, are essential for the cohesiveness of small traditional societies [65].

These relationships are part of the unique dynamics of cultural niche-construction, as humans must adapt to a 'cognitive niche' of knowledge-using and socially interdependent individuals, among whom "cognitive schemas and social emotions that evolved for one domain can be pressed into service for another and assembled into increasingly complex mental structures," [64]. During the 20th century, this cognitive niche began to include global economics as part of the shared emotional experience of populations.

It is certain that more literary cycles of this sort and their correlations with world events will be discovered; showing that in addition to predicting the near future using very recent event data [25], big-data can also be used to understand cultural dynamics on much longer time scales [23,26,38,39,59].

Methods

We extracted literary mood from the Ngram corpus using lists of semantically related terms provided in WordNet Affect [43–45], and we additionally validated this method with two alternative and independent tools, the Linguistic Inquiry and Word Count [46], and the recent "hedonometer" method [47], originally proposed to analyse Twitter data.

We obtained the time series of words frequencies from the Google Books Ngram corpus using Version 2 (released in July 2012). Version 2 includes more than 8 millions books (versus the

Table 1. Correlations between the 11-year moving average on the U.S. misery index and different literary samples, 1929 to 2000.

Literary index	Pearson's r	p (<) that r	std. dev. on r	95% conf. interval on r
WNA misery (all English books)	0.79	0.0005	0.05	0.67–0.87
WNA misery (American books)	0.65	0.0005	0.07	0.51–0.76
LIWC misery	0.55	0.0005	0.08	0.37–0.70
WNA misery (British books)	0.46	0.0005	0.09	0.27–0.62
Hedonometer misery	0.31	0.01	0.12	0.00–0.49
LIWC negative emotions	0.27	0.05	0.11	0.03–0.47
WNA disgust (all English books)	−0.25	0.05	0.13	−0.49–0.17
LIWC sadness	0.23	0.05	0.13	−0.01–0.50

Highest magnitude correlations at top. Standard deviations and confidence intervals on r are estimated by 10,000 bootstraps (random sampling with replacement). Not shown are the literary scores which did not yield statistically significant correlations with economic misery: WNA misery (Fiction books), WNA joy, WNA fear, WNA surprise, WNA anger, WNA sadness, LIWC anxiety, LIWC anger, LIWC affect, and LIWC positive emotions.

doi:10.1371/journal.pone.0083147.t001

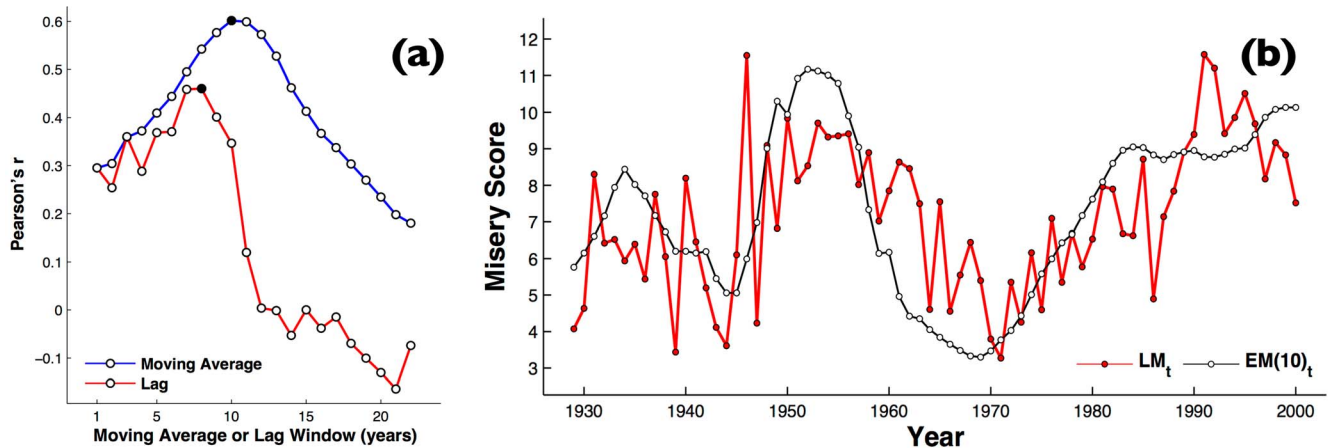


Figure 4. Repeating the analysis on German economic and literary misery. (a) The effect of varying the moving average period, τ , versus as a simple lag, on correlation between for German LIWC and German misery index (b) Time series of the literary misery index for all books calculated through German LIWC (white circles), versus the German economic misery index (red circles). doi:10.1371/journal.pone.0083147.g004

about 5 millions of the previous version) and better OCR results [37]. We considered English-language books, in four distinct corpora (all books in English, fiction books in English, American English books, and British English books).

We used the WordNet Affect (henceforth WNA) text analysis tool which groups synonymous terms into lists related to mood states to perform text analysis on these words after they had been stemmed using Porter's Algorithm [66]. This method is consistent with numerous other text-mining studies such as [22,67]. We considered six distinct main emotions: anger ($N=146$), disgust ($N=30$), fear ($N=92$), joy ($N=224$), sadness ($N=115$), and surprise ($N=41$).

For each stemmed word we collected the amount of 1-gram occurrences (case insensitive) in each year from 1900 to 2000 (both included). Following [38], we excluded data from years after 2000 because books published recently are still being included in the data set, and therefore latest records are incomplete and possibly biased. We normalized the frequency of the words in these word lists, then computed the average normalized frequency per year:

$$w_i = \frac{1}{n} \sum_{i=1}^n \frac{c_i}{c_{the}}, \quad (1)$$

where n is the number of words in the list, c_i is the word count for word i in the list in year t , normalized by c_{the} , which is the count of the most frequent English word, 'the', in year t . We normalized the yearly amount of occurrences using the occurrences, for each year, of the word 'the', rather than by the annual total number of words scanned, to avoid the effect of the influx of data, special characters, etc. that may have changed considerably over the 20th century [40,68].

These normalized frequency scores were then converted to their z -score equivalents as

$$(w_i - \mu_w) / \sigma_w, \quad (2)$$

where μ_w and σ_w are the mean and standard deviation of w_i over the 100 years of the 20th century. We denote the z -score equivalents in year t for the 224 listed 'joy' words and the 115 listed 'sadness' words as J_t and S_t , respectively, and the difference between them, $LM_t = S_t - J_t$ as our literary misery index, LM_t .

We note that normalising by 'the' is not a significantly determinant of our results. Over the twentieth century, frequencies of 'the' exhibit none of the cycling patterns of the emotion words. To confirm that this was not a factor, we re-calculated the literary misery index from All Words, but normalised this time with the yearly frequencies of the word, 'of' rather than 'the', in the denominator of eq.(1). Using these re-normalised values of LM_t we find virtually identical results to those presented in our Results section: a peak in the overall goodness of fit with economic misery at $\tau=10-11$ (Pearson's $r=0.47$ for either), which again is better than the best value for a simple lag ($r=0.42$ at a lag of 7 years), and is much better for the period 1929 to 2000 (best fit at $\tau=10$ years (Pearson's $r=0.72$)).

We also performed the same mood scoring process using different taxonomies of emotional words taken from the Linguistic Inquiry and Word Count (henceforth LIWC), for which the vocabularies of emotions have been evaluated word-by-word by human judges [23]. In particular, we consider the LIWC categories of general affect ($N=917$ terms), anger ($N=184$), anxiety ($N=91$), negative emotions ($N=499$), positive emotions ($N=408$) and sadness ($N=101$). Unlike WNA, LIWC already includes the stems of words together with complete (non-stemmed) words in all of its vocabularies. The analysis was performed as described for WNA (in a non-Porter-stemmed version of the Ngram corpus) and we calculated the literary misery index, LM_t , as the difference between the z -score equivalents in year t for the 101 listed 'sadness' words and for the 408 listed 'positive emotions' words. All the analyses were performed on the main corpus (i.e. all books in English). For the LM_t trends derived from WNA, which are the main focus of our analysis, we used additionally the other corpora (fiction books in English, American English books, and British English books).

The last method applied to English was to calculate the literary misery index ("hedonometer") is described in details in [47]. In this case we extracted from the Ngram corpus the data for 3,686 words which were previously evaluated for their 'happiness content' using Amazon's Mechanical Turk. Their normalised frequencies were then weighted with the values provided in [47]. To compare those results with the previous analysis, we z -scored them and we considered the opposite of the original index (since we are interested in literary 'misery'). This method is conceptually different from WNA and LIWC, because it does not consider

strictly mood-words; additional terms that are assigned positive or negative feeling, such as 'food' (positive) or 'funeral' (negative).

To check the generalisation of the results, we also analysed books written in German. The LIWC also allows an analysis of German text as it provides the same emotional categories in German language [69]. To compute a German literary misery index, we used the German corpus of Google Books, which includes approximately 660,000 books, and the analogous LIWC categories of positive emotions ($N=645$) and sadness ($N=186$). We followed the same methodology as for English books, and for consistency word counts were normalized using the aggregate count of 'die', 'der', 'das' (feminine, masculine, and neuter versions of 'the').

References

- Giguère G, Love BC (2013) Limits in decision making arise from limits in memory retrieval. *Proc Natl Acad Sci USA*, in press (DOI 10.1073/pnas.1219674110)
- Olivola CY, Sagara N (2009) Distributions of observed death tolls govern sensitivity to human fatalities. *Proc Natl Acad Sci USA* 106: 22151–22156.
- Stewart N, Chater N, Brown GDA (2006) Decision by sampling. *Cognitive Psychology* 53:1–26.
- Petersen AM, Tenenbaum J, Havlin S, Stanley HE, Perc M (2012) Languages cool as they expand: Allometric scaling and the decreasing need for new words. *Sci Rep* 2: 943.
- Henrich J (2004) Demography and cultural evolution: Why adaptive cultural processes produced maladaptive losses in Tasmania. *American Antiquity* 69: 197–214.
- Kline M, Boyd R (2010) Population size predicts tool complexity in Oceania. *Proc Roy Soc B* 277: 2559–2564.
- Mesoudi A, Whiten A (2008) The multiple uses of cultural transmission experiments in understanding cultural evolution. *Phil Trans Roy Soc B* 363: 3489–3501.
- Powell A, Shennan SJ, Thomas MG (2009) Late Pleistocene demography and the appearance of modern human behavior. *Science* 324: 1298–1301.
- Woolley AW, Chabris CF, Pentland A, Hashmi N, Malone TW (2010) Evidence for a collective intelligence factor in the performance of human groups. *Science* 330: 686–688.
- Carroll J (1995) Evolution and literary theory. *Human Nature* 6:119–134.
- Pinker P, Nowak MA, Lee JJ (2008) The logic of indirect speech. *Proc Natl Acad Sci USA* 105:833–838.
- Bandelj N (2009) Emotions in economic action and interaction. *Theory and Society* 38: 347–366.
- Bagozzi RP, Gopinath M, Nyer PU (1999) The role of emotions in marketing. *J Acad Marketing Sci* 27: 184–206.
- Bechara A, Damasio AR (2005) The somatic marker hypothesis: A neural theory of economic decision. *Games Econ Behav* 52: 336–372.
- Berezin M (2009) Exploring emotions and the economy. *Theory and Society* 38: 335–346.
- Loewenstein G (2000) Emotions in economic theory and economic behavior. *Am Econ Rev* 90: 426–432.
- Lieberman E, Michel J-P, Jackson J, Tang T, Nowak MA (2007) Quantifying the evolutionary dynamics of language. *Nature* 449: 713–716.
- Berger J, Le Mens G (2009) How adoption speed affects the abandonment of cultural tastes. *Proc Natl Acad Sci USA* 106: 8146–8150.
- Bakshy E, Hofman JM, Mason WA, Watts DJ (2011) Everyone's an influencer. In King I, Neijdl W, Li H, editors, *Proc WSDM11, Hong Kong*.
- Acerbi A, Ghirlanda S, Enquist M (2012) The logic of fashion cycles *PLoS ONE* 7(3): e32541.
- Kloumann IM, Danforth CM, Harris KD, Bliss CA, Dodds PS (2012) Positivity of the English language. *PLoS ONE* 7(1): e29484.
- Lansdall-Welfare T, Lampos V, Cristianini N (2013) Effects of the recession on public mood in the UK. *Proc 21st international conference companion on World Wide Web*: 1221–1226.
- Pennebaker JW (2011) *The Secret Life of Pronouns: What our Words Say About Us*. New York: Bloomsbury Press.
- Gao J, Hu J, Mao X, Perc M (2012) Culturomics meets random fractal theory: Insights into longrange correlations of social and natural phenomena over the past two centuries *J R Soc Interface* 9: 1956–1964.
- Bollen J, Mao H, Zeng X-J (2011) Twitter mood predicts the stock market. *J Comp Sci* 2: 1–8.
- DeWall CN, Pond RS Jr, Campbell WK, Twenge JM (2011) Tuning in to psychological change. *Psychology of Aesthetics, Creativity and the Arts* 5: 200–207.
- Lampos V, Preotiuc-Pietro D, Cohn T (2013) A user-centric model of voting intention from Social Media. In *Proc 51st Annual Meeting of the Association for Computational Linguistics*: 993–1003.
- Twenge J, Campbell KW, Gentile B (2012) Increases in individualistic words and phrases in American books, 1960–2008. *PLoS ONE* 7(7): e40181.
- Choi H, Varian H (2012) Predicting the present with Google Trends. *Economic Record* 88: 2–9.
- Ettredge M, Gerdes J, Karuga G (2005) Using web-based search data to predict macroeconomic statistics. *Communications of the ACM* 48: 87–92.
- Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, et al. (2008) Detecting influenza epidemics using search engine query data. *Nature* 457: 1012–1014.
- Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ (2010) Predicting consumer behavior with web search. *Proc Natl Acad Sci USA* 107:17486–17490.
- Lampos V, Cristianini N (2012) Nowcasting events from the Social Web with statistical learning. *ACM Transactions on Intelligent Systems and Technology* 3: article 72.
- Lazer D, Pentland A, Adamic L, Aral S, Barabasi AL, et al. (2009) Computational social science. *Science* 323: 721–723.
- Saavedra S, Hagerty K, Uzzi B (2011) Synchronicity, instant messaging, and performance among financial traders. *Proc Natl Acad Sci USA* 108: 5296–5301.
- Preis T, Moat HS, Stanley HE, Bishop SR (2012) Quantifying the advantage of looking forward. *Scientific Reports* 2: 350.
- Lin Y, Michel JB, Alden EL, Orwant J, Brockman W, et al. (2012) Syntactic annotations for the Google Books Ngram Corpus. *Proc ACL 2012 System Demonstration*: 169–174.
- Michel J-B, Shen YK, Aiden AP, Veres A, Gray MK, et al. (2011) Quantitative analysis of culture using millions of digitized books. *Science* 331: 176–182.
- Perc M (2012) Evolution of the most common English words and phrases over the centuries. *J R Soc Interface* 9: 3323–3328.
- Acerbi A, Lampos V, Garnett P, Bentley RA (2013) The expression of emotions in 20th century books. *PLoS ONE* 8(3): e59030.
- Okun AM (1981) *Prices and Quantities: A Macroeconomic Analysis* Brookings Institution.
- Ormerod P, Rosewell B, Phelps P (2013) Inflation/unemployment regimes and the instability of the Phillips curve. *Applied Economics* 45: 1519–1531.
- Strapparava C, Valitutti A (2004) Wordnet-affect: an affective extension of wordnet. *Proc 4th International Conference on Language Resources and Evaluation*: 1413–1418.
- Strapparava C, Mihalcea R (2008) Learning to identify emotions in text. *Proc SAC'08*: 1556–1560.
- Miller GA (1995) WordNet: a lexical database for English. *CACM* 38(11): 39–41.
- Pennebaker JW, Booth RJ, Francis ME (2007) *Linguistic Inquiry and Word Count: LIWC2007*. LIWC.net, Austin, Texas.
- Dodds PS, Harris KD, Kloumann IM, Bliss CA, Danforth CM (2011) Temporal patterns of happiness and information in a global-scale social network: Hedonometrics and Twitter. *PLoS ONE* 6: e26752.
- Sumner WG (1959) *Study of the Sociological Importance of Usages, Manners, Customs, Mores and Morals*. Dover Publications.
- Abrahamson E, Eisenman M (2008) Employee-management techniques: Transient fads or trending fashions? *Administrative Science Quarterly* 53: 719–744.
- Turchin P, Nefedov S (2009). *Secular Cycles*. Princeton University Press.
- McKee-Ryan FM, Song Z, Wanberg CR, Kinicki AJ (2005) Psychological and physical well-being during unemployment: A meta-analytic study. *Journal of Applied Psychology* 90: 53–76.
- Murphy GC, Athanasou JA (1999) The effect of unemployment on mental health. *Journal of Occupational and Organizational Psychology* 72: 83–99.
- Steward MJ, Makwarimba E, Reutter LI, Veenstra G, Raphael D, et al. (2009) Poverty, sense of belonging, and experiences of social isolation. *Journal of Poverty* 13: 173–195.
- Paul KI, Moser K (2009) Unemployment impairs mental health: Meta-analysis. *Journal of Vocational Behavior* 74: 264–282.

Acknowledgments

We thank J. Pennebaker for providing the lists of emotion-related words from LIWC-2007. We thank J. Carrol and B. Wuyts for comments.

Author Contributions

Conceived and designed the experiments: RAB AA VL. Analyzed the data: RAB AA VL. Wrote the paper: RAB AA VL PO.

55. Kivimäki M, Vahtera J, Pentti J, Ferrie JE (2000) Factors underlying the effect of organisational downsizing on health of employees: a longitudinal cohort study. *BMJ* 320:971–975.
56. Sullivan D, von Wachter T (2009) Job displacement and mortality *Quarterly Journal of Economics* 124: 1265–1306.
57. Kivimäki M, Vahtera J, Virtanen M, Elovainio M, Pentti J, et al. (2003) Temporary employment and risk of overall and cause-specific mortality. *Am J Epidemiol* 158: 663–668.
58. van Lenthe FJ, Borrell LN, Costa G, Diez Roux AV, Kauppinen TM, et al. (2005) Neighbourhood unemployment and all cause mortality: A comparison of six countries. *Journal of Epidemiological Community Health* 59: 231–237.
59. Hughes JM, Foti NJ, Krakauer DC, Rockmore DN (2012) Quantitative patterns of stylistic influence in the evolution of literature. *Proc Natl Acad Sci USA* 109: 7682–7686.
60. Quoidbach J, Gilbert DT, Wilson TD (2013) The end of history illusion. *Science* 339:96–98.
61. Elder GH (1974) *Children of the Great Depression*. Westview Press.
62. Oishi S, Graham J, Kesebir S, Galinha IC (2013) Concepts of happiness across time and cultures. *Pers Soc Psych Bul* 39: 559–577.
63. Mullins DA, Whitehouse H, Atkinson QD (2013) The role of writing and recordkeeping in the cultural evolution of human cooperation. *J Econ Behav Org* 90S: S141–S151.
64. Pinker S (2010) The cognitive niche: Coevolution of intelligence, sociality, and language. *Proc Natl Acad Sci USA* 107: 8993–8999.
65. Marshall L (1957) The kin terminology of the !Kung Bushmen. *Africa* 27: 1–25.
66. Porter MF (1980) An algorithm for suffix stripping. *Program* 14: 130–137.
67. Calvo RA, D'Mello S (2010) Affect detection: An interdisciplinary review of models, methods, and their applications. *TAC* 1(1): 18–37.
68. Bentley RA, Garnett P, O'Brien MJ, Brock WA (2012) Word diffusion and climate science. *PLoS ONE* 7(11): e47966.
69. Wolf M, Horn A, Mehl M, Huang S, Pennebaker JW, et al. (2008) Computergestützte quantitative Textanalyse: Äquivalenz und Robustheit der deutschen Version des Linguistic Inquiry Word Count. *Diagnostica* 2: 85–98.
70. Romer C (1986) Spurious volatility in historical unemployment data. *J Pol Econ* 94(1): 1–37.
71. Coen RM (1973) Labor force and unemployment in the 1920's and 1930's: A re-examination based on postwar experience. *Rev Econ Stat* 55(1): 46–55.
72. Mitchell BR (2007) *International Historical Statistics: Europe, 1750–2005*. New York: Palgrave Macmillan.
73. Maddison A (2003) *The World Economy: Historical Statistics*. Paris: Organization for Economic Co-operation and Development.